Arctic Network
Inventory & Monitoring Program

U.S. Department of the Interior

Data Management
Standard Operating Procedure
NPS/ARCN/DMSOP-2013-01

Connecting to the Inventory and Monitoring Climate Database using R

Summary

This Standard Operating Procedure documents the process for accessing information from the Inventory and Monitoring Climate Database for statistical and graphical analysis using the R statistical package and RODBC database connectivity library. Example scripts are provided.

Contents

Summary 1
Contents 1
Introduction 1
Creating an Open Database Connectivity (OCBC) database connection 1
Installing RODBC package 4
Connecting to the Climate Database using R and RODBC 4
A sample session: Getting a list of climate station dataset versions 4
A sample session: Plotting data from a climate station data version 5
Where to go from here 7
About This Standard Operating Procedure 8
Revision History 8

Introduction

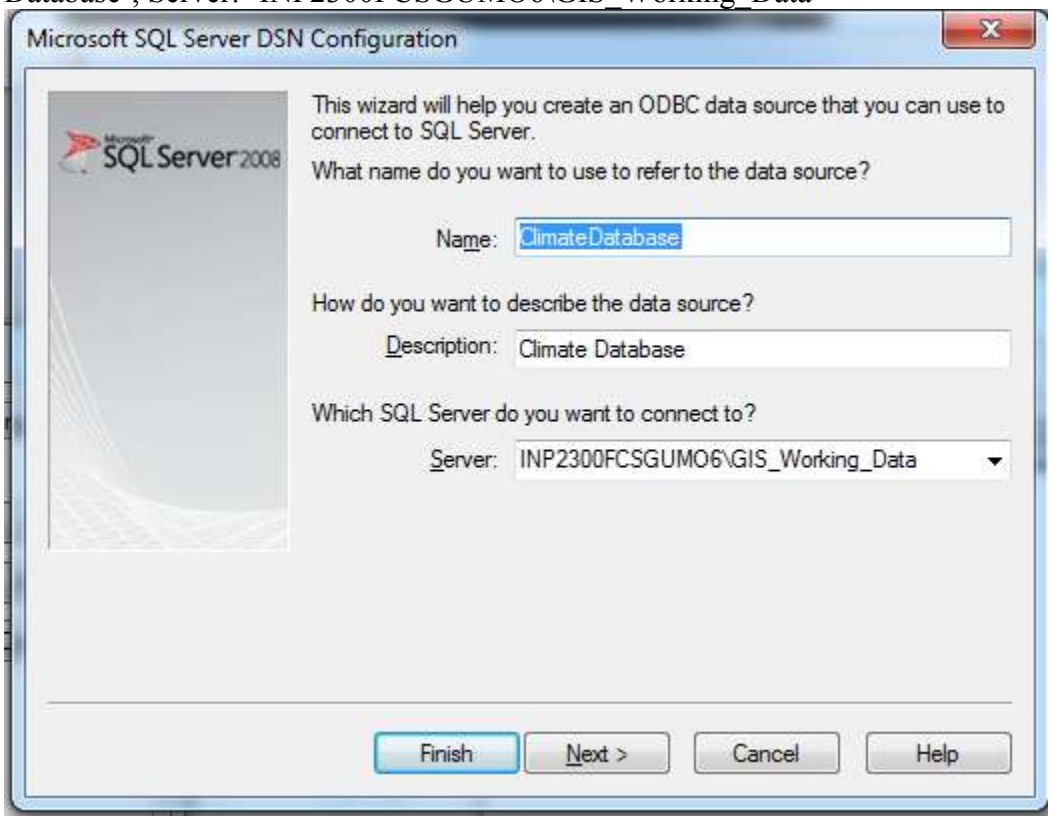
The National Park Service Inventory and Monitoring program maintains a climate database for storing and retrieving climate data from a variety of data sources. Information about the climate database can be found at https://irma.nps.gov/App/Reference/Profile/2167254. This SOP describes the general process of connecting to the climate database using the R statistical package from a computer running Windows 7 operating system.

Creating an Open Database Connectivity (OCBC) database connection

Steps:

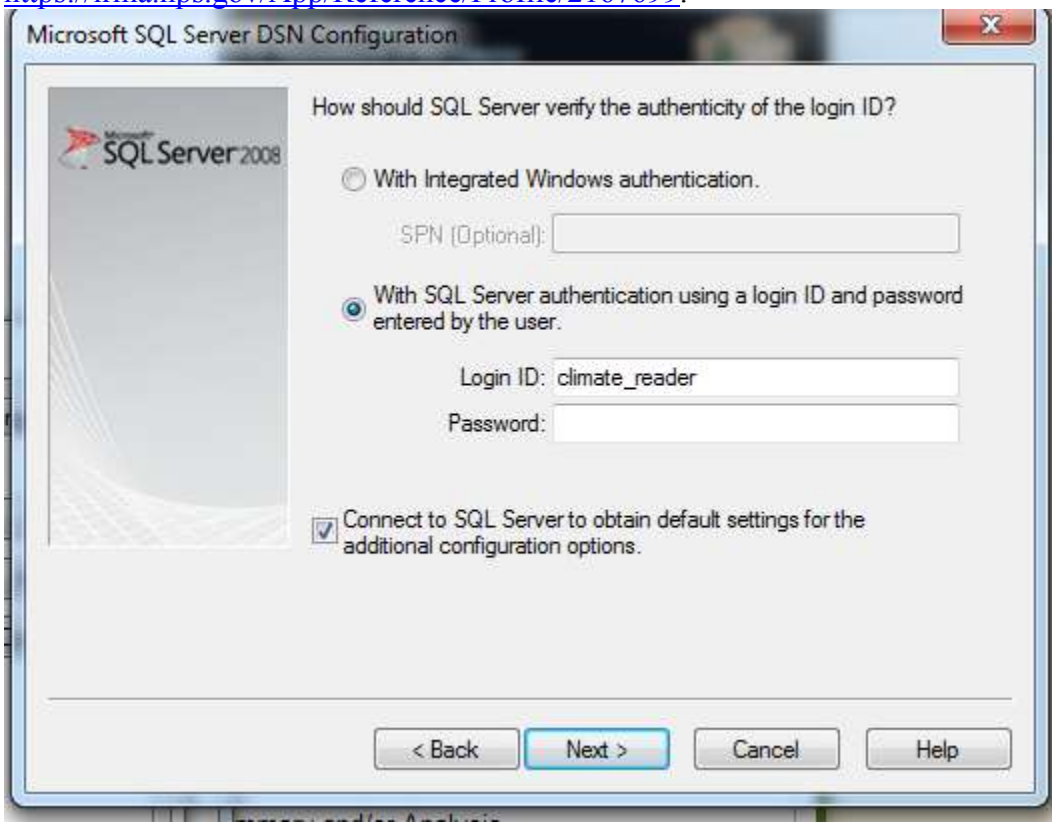
1. From your Start Menu search text box type in 'ODBC' and select 'Data Sources (ODBC)'
2. Click 'Add'

3. Select 'SQL Server Native Client 10.0' or similar. *If you don't see a SQL Server driver then search for one on the Microsoft website and install it (these URLs change so frequently that including a link here would be pointless)*
4. Click Finish
5. Fill in the fields as follows: Name: 'ClimateDatabase', Description: 'Climate Database', Server: 'INP2300FCSGUMO6\GIS_Working_Data'

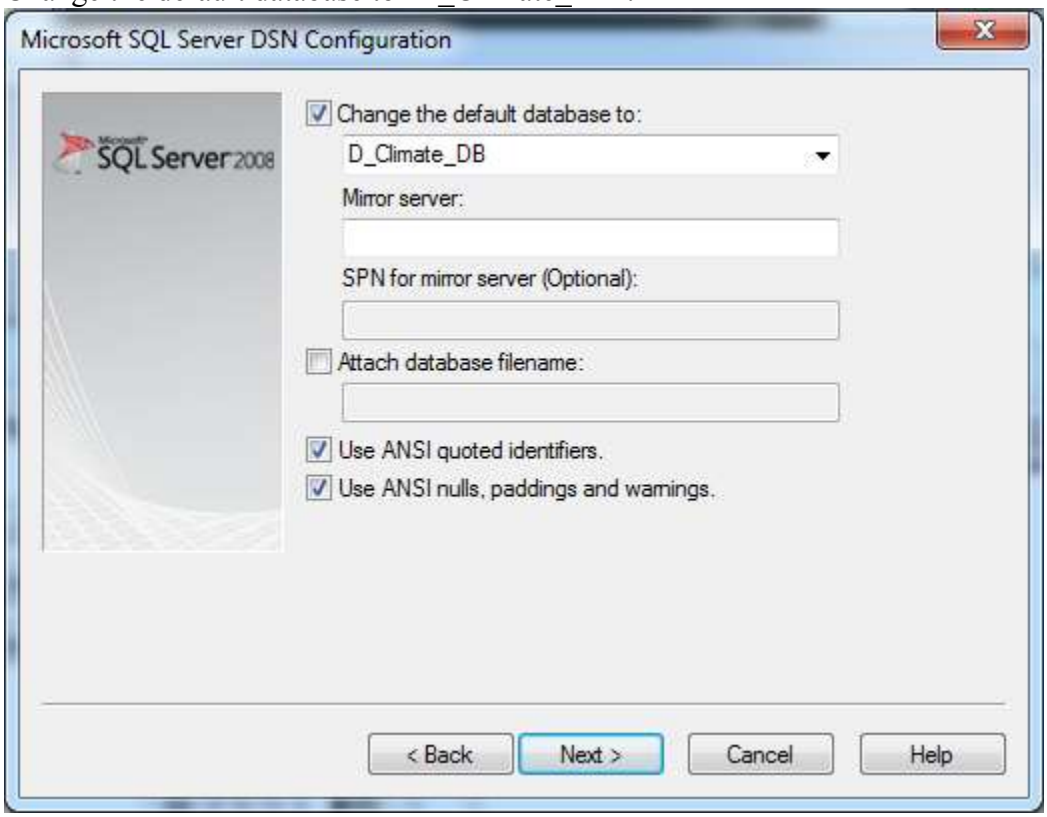


6. Click Next
7. The next page asks how you would like to authenticate yourself to the database. Select 'SQL Server authentication' and enter the login 'climate_reader' and password for that account. The password is available in the SOP titled 'Connecting to the I&M Enterprise Climate Database for the Purpose of Data Retrieval, Summary and/or Analysis' available at

https://irma.nps.gov/App/Reference/Profile/2167699.



8. Click Next.
9. Change the default database to 'D_Climate_DB'.



10. There are no further edits needed so click your way through the rest of the wizard keeping the default values and click Finish when through.
11. It's a good idea to test the connection by clicking the 'Test Data Source' button to make sure the data source works correctly.
12. Click OK.

Installing RODBC package

This SOP will assume you have installed R and whatever Graphical User Interface (GUI) you prefer. In order for R to communicate with the SQL Server hosting the climate database you will need to download and install the RODBC package. You can obtain RODBC at http://cran.r-project.org/web/packages/RODBC/index.html and install it according to the R documentation (http://cran.r-project.org/doc/manuals/R-admin.html#Windows-packages) or by following the instructions provided by your R GUI.

Connecting to the Climate Database using R and RODBC

Connect to the database by importing the RODBC package:

```
#import the R ODBC library so we can connect to sql server
library(RODBC)
```

Connect to the database using the climate_reader SQL Server account. "ClimateDatabase" is the ODBC data source name. Create this data source if you haven't done so already.

```
DBConnection <-odbcConnect("ClimateDatabase",
uid="climate_reader", pwd="password")
```

A sample session: Getting a list of climate station dataset versions

The following script imports the RODBC library, connects to the database via the ClimateDatabase ODBC connection and queries the database for information about climate station data versions in three parks. This query could be used to choose a version for further data analysis.

```
#import the R ODBC library so we can connect to sql server
library(RODBC)

#connect to the database using the climate_reader SQL
Server account.  "ClimateDatabase" is the ODBC data source
name.  Create this data source if you haven't done so
already.
```

```
DBConnection = odbcConnect("ClimateDatabase",
uid="climate_reader", pwd="password")

#store our query in a variable
Sql = "SELECT     StationsInDatabase.UnitList,
StationsInDatabase.StationName, VersionInfo.ParameterCode,
VersionInfo.Unit, VersionInfo.versionID,
VersionInfo.MinDate, VersionInfo.MaxDate,
VersionInfo.NumDaysDiff, VersionInfo.NumObs,
VersionInfo.PercentObsPresent,
StationsInDatabase.StationCode,
StationLocations.Latitude_deg,
StationLocations.Longitude_deg FROM StationsInDatabase
INNER JOIN VersionInfo ON StationsInDatabase.StationCode =
VersionInfo.StationCode LEFT OUTER JOIN StationLocations ON
StationsInDatabase.StationCode =
StationLocations.StationCode WHERE
(StationsInDatabase.UnitList IN ('DENA', 'YUCH', 'WRST'))
ORDER BY StationsInDatabase.UnitList,
StationsInDatabase.StationName, VersionInfo.ParameterCode"

#store our data in a variable
VersionsList = sqlQuery(DBConnection,Sql)

#display the list of climate data versions
VersionsList

#close the database connection (good housekeeping practice)
close(DBConnection)
```

A sample session: Plotting data from a climate station data version

The following script imports the RODBC library, connects to the database via the ClimateDatabase ODBC connection and queries the database for minimum daily temperature at the Upper Yukon Charley River Alaska station (VersionID: 61857). The data is stored in a variable called Version which can then be submitted to a plot function to generate a graph.

```
#import the R ODBC library so we can connect to sql server
library(RODBC)

#connect to the database using the climate_reader SQL
Server account.  "ClimateDatabase" is the ODBC data source
name.  Create this data source if you haven't done so
already.
DBConnection = odbcConnect("ClimateDatabase",
uid="climate_reader", pwd="password")
```

```
#set VersionID to the versionid of the data you would like
to see.
VersionID = 61857 #upper yukon charley station

#make a variable to hold our sql query
Sql = paste("SELECT Agency, SourceCode, StationCode,
StationName, ObsDate, ParameterCode, ParameterValue,
ParameterUnit, DataQualityFlagList, DataMeasureFlagList,
DataProcessingFlagList, VersionID, ObservationID,
IsRawData, SiteID, StationID, CDB_Code FROM
vBasicSelectionWithFlags WHERE     (VersionID =  ",
VersionID, ") ORDER BY ObsDate")

#get the data into a variable
Version <- sqlQuery(DBConnection,Sql)

#create a simple graph
plot(Version$ObsDate, Version$ParameterValue,
type="o",col="black",main=Version$StationName[1],
sub=Version$ParameterCode[1],xlab="Date", ylab="Minimum
Temperature (F)")

#close the database connection (good housekeeping practice)
close(DBConnection)
```

The script above results in the following graph (Figure 1).

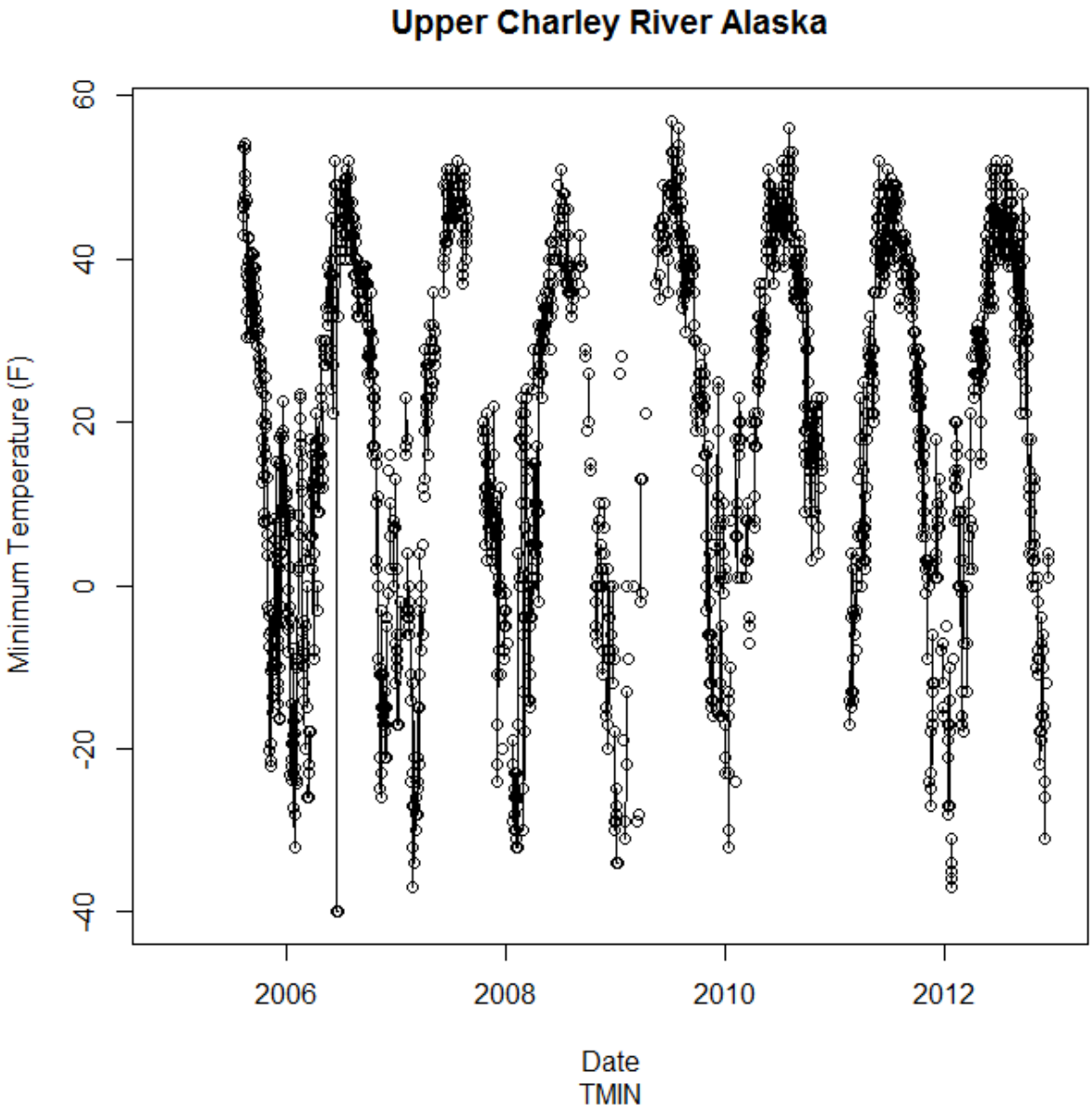


Figure 1. Example minimum temperature graph.

Where to go from here

This SOP is very limited in scope and only pertains to setting up an environment for exploration of the I&M Climate Database using R. The Climate Database designers have set up numerous queries that are immediately available for use using the example scripts above as a starting point. They have done the hard work for you.

From here you are strongly encouraged to read *Connecting to the I&M Enterprise Climate Database for the Purpose of Data Retrieval, Summary and/or Analysis*. The authors present numerous 'ready baked' queries and functions such as annual summaries,

growing degree days, monthly summaries, etc. One merely has to copy and paste the query from the SOP into the scripts from above.

About This Standard Operating Procedure

Version: 1.0
Status: Draft
Publication Date: January 25, 2013
Author(s): Scott D. Miller, Data Manager, Arctic and Central Alaska Networks Inventory & Monitoring Program.
Abstract: This Standard Operating Procedure documents the process for accessing information from the Inventory and Monitoring Climate Database for statistical and graphical analysis using the R statistical package and RODBC database connectivity library. Example scripts are provided.
Suggested Citation: NPS-ARCN (2013). Connecting to the Inventory and Monitoring Climate Database using R. Version 1.00, Arctic Network-Inventory and Monitoring Program, National Park Service. Fairbanks, Alaska.

Revision History

Version	Version Date	Revised By	Changes
1.0	2010-01-11	S. Miller	Original

This table reflects changes to this document. Version numbers will be incremented by one (e.g., Version 1.3 to Version 2.0) each time there is a significant change in the process and/or changes are made that affect the interpretation of the data. Version numbers will be incremented after the decimal (e.g., Version 1.6 to Version 1.7…1.10….1.21) when there are changes to grammar, spelling, or formatting, or minor modifications in the process that do not affect the interpretation of data.